Summary report: ESFRI Workshop on the Future of Research Infrastructures in the European Research Area

6-8 November 2019, Los Cancajos, La Palma, Canary Islands

Session 1. The role of Research Infrastructures and ESFRI in EOSC

Session Chair: Mirjam van Daalen

Session Objectives

(As formulated in the program)

This session will take stock of the current state of implementation of the EOSC and the links of Research Infrastructures with this process. Ongoing activities will be discussed, and the remaining challenges explored. The discussion will in particular focus on how ESFRI follows up the EOSC development and how it can contribute to its ongoing or future activities. ESFRI EOSC Task Force as an important advisory tool will have a central role in the discussion, and in the preparation of a future oriented ESFRI position related to EOSC. The importance of ESFRI RI clusters, their approaches and their achievements related to EOSC implementation will flow into the debate. An important issue will be communication to the different EOSC bodies, and the question how to effectively spread ESFRI inputs to the EOSC activities to create biggest impact on its implementation. The event will also offer an opportunity to consider new topics for a potential dedicated ESFRI Exchange of Experience workshop on related matters.

Session Presentations

- Introduction – 15', Mirjam van Daalen, Chair of ESFRI Task Force on EOSC *(download the presentation)*
- Interplay of horizontal e-Infrastructures and thematic RIs – 15', Ludek Matyska, Director, Institute of Computer Science, Masaryk University *(download the presentation)*
- The e-IRG National Nodes Paper – 15', Sverker Holmgren, Uppsala University, Swedish Delegate to e-IRG *(download the presentation)*
- Getting ready for EOSC: a research infrastructure perspective – 15', Nicholas Pade, Executive Director, EMBRC ERIC *(download the presentation)*

Round-tables introduction

The EOSC is a complex subject and the questions posed in the session, despite appearing straightforward at first sight, are inter-connected in many ways. Participants were actively engaged in discussions and gave full consideration to the issues as a whole, and consequently the table responses reflect the complexity and inter-connectivity of the topic. In analysing the responses, an attempt has been made to focus the summary in order to address both the relevant question and the objectives as a whole, however, we must also take care to avoid over-simplification of this challenging and multi-faceted subject.

Round-table discussion focus

There were around 100 participants across Europe, who were grouped in tables of around 10 persons in each table. Participants were asked to engage in round table discussions focussed on the following 5 questions. Note not all questions were discussed by all tables.

1. **What is the EOSC according to you as a user and/or provider?**

Discussion of this question demonstrated that a shared understanding of what the EOSC is (or will be) is still evolving in the RI community. Some clear points surrounding the definition of the EOSC by this audience could be drawn, including its aims (open and reproducible science, FAIR data and services), collaborative elements (horizontal and thematic) and the understanding that there would be user access point(s) with an underlying interoperable infrastructure. Priority areas for further consideration emerged, including the approach (implementation and beyond), data, and accessibility (both tools and services).

2. **Which benefits should the EOSC bring to the users?**

Within this community, there are a number of clear, shared expectations regarding the benefits that the EOSC and its methodological/regulatory framework can deliver. These included primary scientific advantages, such as enhanced data and service connections, ability to address societal challenges, and improved tools for RIs, but also political, social and cultural advantages, for example improved political decision making capabilities, increased societal awareness and leading a change in culture towards open science. Participants also identified a number of qualifying factors, such as the need for transparency, high data quality, research acknowledgement/credit, and training.

3. **Which kind of EOSC users would you define?**

Participants identified a range of potential users of the EOSC, noting that successful engagement and user access is contingent on the effective implementation of supporting services. Different roles of user and provider were also identified. User-groups included RIs, broad scientific community, computation scientists, industry and machine/service users, policy-makers, and additionally citizens as end users. Participants also noted user limitations/conditions which may vary depending on, for example, user group (public sector vs industry), sensitive data restrictions, and embargo periods.

4. **For EOSC there needs to be a change of culture to open research data according to the FAIR data principles: how can we tackle this?**

This question provoked some of the most extensive and detailed dialogue amongst participants, illustrating the challenges in balancing some of the key issues in this complex area. The main directions identified however to achieve this cultural change are education and training, incentives and reward mechanisms, and understanding and funding of associated costs of open science. There must also be further consideration given to the ease of sharing and sharing cultures across disciplines or infrastructures, the type of data, the type of data producer and data user. Participants clearly recognised advantages of the EOSC, including the many benefits of open science, and the 'added value' for RIs. There was significant discussion of open vs FAIR concepts, business models, access limitations, restricted/sensitive data, the importance of data quality, original researcher credit, and training. The RI community was identified as a key player in leading the change of culture and driving data quality and access.

5. **Other topics for workshops**

Participants made many valuable suggestions for future discussion and workshop needs, noting significant overlap between many of them. Participants also identified that there is likely to be an issue of relevance and timeliness of sessions aligning with the EOSC implementation timeline. Broad themes included 'The interactions and dynamics between the different EOSC stakeholders and RIs', 'Fostering interdisciplinary and multidisciplinary collaborations', 'Data', 'Exchange of experience', and 'Funding'.

A full summary and evaluation of responses is provided below.

Draft Recommendations

Informed by the outcomes of the detailed discussions of the workshop (summarised and evaluated below), it is possible to identify a number of high level actions for further consideration by the ESFRI Task Force on EOSC and the discussion with EOSC key stakeholders. To a large extent, the activities highlighted below represent a

continuance of ESFRI's current and ongoing engagement with the EOSC, though they are presented here having been reinforced by the outcomes of the La Palma session.

1. Identify and define key stakeholders and their roles in both understanding and informing the ESFRI/ESFRI RIs position on EOSC,
2. Develop opportunities and mechanisms which facilitate exchange of experience and the building of a common understanding among ESFRI, RIs and EOSC stakeholders,
3. Work with RIs and other relevant stakeholders to develop a position which reflects the unique perspective of this important sector of the research community (i.e. the RIs), using in particular the EOSC cluster projects
4. Engage with EOSC governance bodies through appropriate mechanisms such as the ESFRI Task Force on EOSC and the EOSC clusters to inform and influence the shape of the EOSC,
5. Develop a communication and feedback loop with RIs to provide confidence and reassurance in the sector that their contributions are being taken into account,
6. Recognise that the development of the EOSC is operating according to demanding timeframes and engage effectively at appropriate points in the implementation timeline
7. Develop a long-term view of ESFRI's role in engaging with the EOSC post-implementation.

Discussion summary and evaluation

Q1: What is the EOSC according to you as a user and/or provider?

This fundamental question resulted in a range of EOSC key elements from participants, still not giving a clear answer. It also raised a number of unanswered (and unanswerable at this time) questions, such as the top-down (policy/funders-driven) vs. bottom-up (user-driven) direction on the future of EOSC and the interplay/balance between national and thematic approaches. This response suggests that there is not yet a shared understanding of what the EOSC is amongst the RI community. Yet, there are some points that can be drawn:

- **Definition – RI point of view**

    - Although still abstract, EOSC can be understood as a framework and an evolving process, aiming at open and reproducible science, leading to FAIR (Findable, Accessible, Interoperable, Reusable) data and services, including a set of rules and policies.

    - Collaboration between horizontal (generic) and thematic (domain-specific) is an inherent part of EOSC and its future success. Cross-thematic activities are also welcome. EOSC cluster projects have a key role in both of these points while they are running.

    - EOSC may have both 'visible' access point(s) for the end user (e.g. portal(s), marketplace(s)), and 'invisible' (transparent to the users) underlying interoperable infrastructures (including machine-to-machine interfaces)

In addition, there were, some key features of discussions which begin to emerge as broad priority areas for this group.

- **Approach – implementation and beyond**

    - Implementation should be user/community-driven. RIs should be actively engaged in shaping the future EOSC. IT experts should be also included in this.

    - Although a full requirements capture leading to the EOSC design is lengthy and difficult at this stage, RI requirements for EOSC must be documented and appropriately applied over the course of its implementation

    - Business models for research and commercial services use need to be defined

    - There is a need for strong coordination across EU (thematic) and country-based levels.

    - A better understanding of commonalities across disciplines will be essential. We need to have connections across domains. The role of mediators who interpret the data across domains is key for such interdisciplinary approaches.

- New EOSC tools/technologies will need to be well-federated to/integrated with the existing ones. This must be done with a high regard for data preservation
● **Data**
    - The data to be available within the EOSC has yet to be defined. The offer should be broad and data should be useful and accessible, but limits/restrictions need to be agreed. Data in different fields will have a different lifetime.
    - Raw RI data is unlikely to be useful outside of the RI.
    - RIs are both users (of EOSC) and data providers (for EOSC) at the same time (i.e. prosumers).

● **Accessibility (tools and services)**
    - Agreement to the principle that data should be <u>useful</u> and <u>accessible in a fair way</u> to users outside the specific RI. Suggestions on how to achieve this included provisioning of processes, methodologies and tools (as data), and availability of supporting services (e.g. mediators who support data interpretation for different communities or users).
    - Delivery of EOSC services/tools in advanced RIs (with a set of existing tools) should be carefully considered to be able to federate and use both new and existing services.
    - Note that delivering high levels of accessibility to EOSC data and services (including RI ones) in a sustainable way are likely to require significant resources (e.g. trained experts/mediators to facilitate use of data from different fields, computation scientists to adapt/improve services/ algorithms)

Q2: Which benefits should EOSC bring for the users?

Amongst this group of participants, table responses demonstrate that there are a number of clear, shared expectations regarding the benefits the EOSC and its methodological/regulatory framework deliver.

At this stage, expectations remain relatively high level but also focussed on primary scientific advantages, including:

● **Enhanced data and services interconnections**, leading to;
    - improved interdisciplinary capabilities
    - amplification of researchers outputs' impact
    - more effective use of resources - original funding investments leveraged across multiple domains, reduced duplication of funding
    - reuse of old/existing data for new results
    - improved collaboration opportunities and research outputs including interdisciplinary research (e.g. by combining the air pollution with biological/health data)

● **Improved tools for RIs**
    - faster and easier findability and access to data and services
    - providing services that single infrastructures cannot
    - common metadata schema for all data/services with a minimum quality
    - cloud storage supporting FAIR data
    - longer term preservation of data (also public asset)
    - utilising EOSC workflows to combine data
    - supporting definition of new experiments more efficiently

Some discussions however went into a higher level and identified further broader benefits (e.g. cultural and socioeconomic), including:
● **Change of researchers' culture towards openness and sharing** through the use of EOSC
● **Improved decision making capabilities**; recognising the need for considering digestion of data to the political sphere
● **Increased societal awareness/perception of research/science and researchers/scientists**
● **Increased level of European competitiveness**

Additionally, a number of other considerations qualifying the above benefits were identified:
● transparency is vital to ensure trust in the EOSC; essential development of 'trust tools' to guarantee data quality, trust and benefits across users and providers in a transparent way (potential use of *blockchain* technology to support this)
● Improve traceability and research credit/acknowledgement/reward to each contributor of data/services/ funding (e.g. researcher, provider, funder)
● training needs must be taken into account
● added value to RIs must be ensured

Participants identified a range of potential users of the EOSC, noting that successful engagement and user access is contingent on the effective implementation of supporting services:

- RIs (noting that RIs are both users and providers of the EOSC)
- Broad scientific community, including individual researchers (end users)
- Computation scientists/developers (who use data to develop improved models, algorithms, tools)
- Private sector/industry/commerce
- Not only 'human' users – machine/services also users

Participants also noted potential limitations/conditions for example;
- different business models for different user categories:
  ▪ academic/RI users vs industry access
  ▪ access by international users based outside EU/nationally funded organisations
- no anonymous access (especially for sensitive/personal/health data)
- qualification of users (model of driving licence)
- embargo periods to enable first use of data by researchers generating it
- sensitive data restrictions

Two tables noted that policy makers and citizens can ultimately become also beneficiaries/end users.

Q4: For EOSC there needs to be a change of culture to open research data according to the FAIR data principles: how can we tackle this?

This question provoked some of the most extensive and detailed dialogue amongst participants, illustrating the challenges in balancing some of the key issues in this complex area. The main directions however to achieve this cultural change are the following:

● Education and training on open science and data sharing at early stages i.e. universities, especially before PhD level. Already early career researchers are more open towards sharing.

● Incentives and reward mechanisms for opening data and data sharing (are key for the change of culture), along with policies from funding agencies (carrot and stick approach), mainly for public funding. For private funding this may be different, i.e. the way research is funded may influence the way data is managed/shared (public funding vs private/industry funding).

● There is a cost of making data open and FAIR (including data/metadata quality and reliability, and FAIR maturity level), which needs to be understood and included in funding proposals.

● There may be differences on the ease of sharing and sharing cultures across disciplines or infrastructures, depending on the type of data (e.g. raw vs. processed, and sensitive vs. non-sensitive), but also the type of data producers (e.g. individual researchers vs. research infrastructures-institutions). For some of the

latter the culture of openness may be more easily-broadly accepted. Research Infrastructures can take a leading role in this. Still, in some of the above cases, the change of culture should be gradual/progressive.

There is clear recognition of the advantages that the EOSC will bring:

- Support for concept of 'open science', enabling access and data sharing across multidisciplinary domains, positioning the EU research community to address systemic and complex societal challenges
- 'Added value' for RIs, recognising that many have yet to develop multidisciplinary coordination, complementarity of data, standardisation, visibility, safety, traceability of data

The fundamental principle which underpins the EOSC – open research data to facilitate open science – does represent a deep cultural shift from existing practices for both user and provider communities. A number of groups identified that this would particularly be the case in academia, the public sector and at the individual researcher level.

Discussions identified that a successful implementation of the EOSC will feature a high regard for the respect, confidence and satisfaction of an engaged community. This will need to take account of:

- the understanding that data generated through use of public funds should be included in the EOSC, but should appreciate limitations on open data concept
    - clear definition of the open science concept
    - open data vs FAIR
    - timeliness of access – embargo periods should be normal accepted practice
    - usability of raw vs processed data (e.g. from RIs)
    - industry and international access, noting potential for misuse of data
    - restricted access to sensitive/personal/health data and implications of GDPR

- the importance of promoting the advantages and new opportunities that the EOSC will offer plus incentivising 'FAIR' and incentivising RIs for collaboration

- the need to recognise/credit original research, noting impact on careers/progression opportunities
    - Code of practice in sharing/using data to support excellent science
    - Use of mandates and incentives (noting that responsibility and resources for this need to be defined)

- How to ensure data quality and reliability
    - Standards for data reliability; procedures and resources to support 'FAIR data maturity labels'
    - Quality control (minimum data quality thresholds, analytical tools, quality controls in each RI)
    - Reproducibility
    - Sustainability and data preservation
    - Funding – these measures are all costly. Who pays?
- Training
    - Early stage research training re. metadata importance and cultural change training (noting also that there may already be a different generational approach towards open data)

One table also identified a role for RIs, whereby specific RIs who have specific competencies in their fields need to work as drivers for the quality of data and access management

Q5: Other topics for workshops

Participants made many valuable suggestions for future discussion and workshop needs. An attempt has been made to group these into broad thematic areas for consideration, though noting significant overlap between many of them. Note also that there is likely to be an issue of relevance and timeliness of sessions aligning with the EOSC implementation timeline.

The most important topic across the majority of tables was the following:

● **The interactions and dynamics between the different EOSC stakeholders and RIs, namely:**
  - Horizontal (generic) e-Infrastructures vs. thematic (domain-specific) RIs, considering that RIs also have e-Infrastructure/data infrastructure built into them. Where is the meeting point and interfaces between the EOSC minimum viable product and the thematic services? How will EOSC gain better understanding of RIs needs?
  - National vs. EU/thematic approaches (including national components of RIs and related funding) and aligning national initiatives with EOSC
  - Related governance aspects, valid from the initial steps to the longer term

Other important areas included:

● **Fostering interdisciplinary and multidisciplinary collaborations**
  - How to foster communication and collaborations across disciplines
  - How to foster interoperability of data/services/tools across disciplines
  - Discussion of different fields for interdisciplinary connections, including;
    ▪ Understanding interdisciplinary potential in the achievement of political goals; how can RIs contribute to EU missions?
    ▪ (meta)data requirements to facilitate use in other disciplines (overlap with data)

● **Data (note different data discussion needs along EOSC implementation timeline)**
  - Data requirements to facilitate interdisciplinary use (overlap with above)
  - Definition of key requirements for data flows to be ready for EOSC
  - How do RIs get data into the EOSC
  - Data access and quality control
  - Who owns data and who is responsible for quality?
  - Data interpretation - understanding technical specialist requirements; new IT professions to support data interpretation, sustainability and career competitiveness
  - Sustainability of data, data preservation
  - What is needed for full openness of data? How to ensure funding for this and how can EOSC facilitate?
  - How fair is FAIR?
  - Sensitive and restricted data
  - Access for different user groups (research vs industry)
  - AI for interdisciplinary research through big data

● **Exchange of experience**
  - "Concept of EOSC" for policy makers and RI managers
  - How will EOSC gain better understanding of RIs
  - Examples from projects on first exchange with EOSC
  - Exchange of experience between RIs and researchers
  -
● **Funding**
  - Coordination and funding at national level
  - Ensuring data quality, incentivising access will be costly. Who will pay?